Information mining of customers preferences for product specifications determination using big sales data

Available online at www.sciencedirect.com

ScienceDirect

Procedia CIRP 109 (2022) 101–106

www.elsevier.com/locate/procedia

32nd CIRP Design Conference

Information mining of customers preferences for product specifications determination using big sales data

Jian Zhang[a,b,*], Peihuang Lin[a], Alessandro Simeone[c]

[a]*Intelligent Manufacturing Key Laboratory of Ministry of Education, Shantou University, Shantou, 515063 China*
[b]*Shantou Institute for Light Industrial Equipment Research, Shantou University, Shantou, 515063 China*
[c]*Department of Management and Production Engineering, Politecnico di Torino, 10129 Turin, Italy*

** Corresponding author. Tel.: +86-13502724586; fax: +86-0754-8251-8900. E-mail address: jianzhang@stu.edu.cn*

Abstract

Product competitiveness is highly influenced by its related design specifications. Information retrieval of customers preferences for the specification determination is essential to product design and development. Big sales data is an emerging resource for mining customers preferences on product specifications. In this work, information entropy is used for customers preferences information quantification on product specifications firstly. Then, a method of information mining for customers preferences estimation is developed by using big sales data. On this basis, a density-based clustering analysis is carried out on customers preferences as a decision support tool for the determination and selection of product design specifications. A case study related to electric bicycle specifications determination using big sales data is reported to illustrate and validate the proposed method.

Keywords: Product design; Customers preferences; Product specification; Information entropy; Big sales data; Density-based clustering

1. Introduction

In the past decades, advanced technologies (e.g., network, information and intelligence technologies) have enabled the design and development of competitive products along with manufacturers technical abilities. However, such sophisticated technologies increased competition among manufacturers as products features and lifecycle performance information can be easily obtained by customers for comparison. Consequently, obtaining customers preferences informations of products quickly and accurately is essential for designing competitive products [1,2].

For a given product category, a set of specifications consisting of metrics and values can precisely describe customers preferences of products [3]. A better understanding of customers both conscious and sub-conscious preferences on product specifications is critical to the specifications determination for competitive product design and development [4].

Existing methods on specifications determination usually include two steps: (1) analysing customers needs, and (2) determining product specifications based on customers needs. [2,3].

In order to better analyse and evalaute the customers needs, a number of methods are available in literature for customer needs classification into different levels and categories based on the strength of customers demands. In the hierarchical theory of needs [5], people needs are divided into 5 different levels ranging from the most basic to the most advanced. Such theory has been applied to classify customers products needs into various levels for supporting design decisions. With reference to product quality features and customers satisfaction, the Kano model aims at classifying customers needs into 5 categories based on the degree to which they are likely to satisfy customers [6,7]. Both hierarchical theory of needs and Kano model are used for qualitative analysis of customer needs. To explore customers perceptual sensibility, relevant literature provides a variety of methods for determining the importance of customers product preferences, such as Kanse engineering method, Analytical Hierarchy Process (AHP) method, and data mining methods for sentiment analysis of customers product preferences using online customers product reviews data [1,7].

10.1016/j.procir.2022.05.221

Quality function deployment (QFD) method is usually employed for mapping customers needs of products into design specifications [8,9]. To better satisfy various customers needs, group preference aggregation methods can also be used for design specification determinations [10,11].

Although existing methods have been widely used in recent years due to their strong applicability and good interpretability, the following limitations of those methods have been observed:

- The used datasets are not big enough as only a limited number of customers can participate in the surveys, and only part of the customers provides useful feedback or comments on the product. This results in an insufficient reflection of customers product preferences in the marketplace.
- Customers preferences on specification combinations are not well considered as few studies were found on investigating specification correlations originated from customers preferences on their combinations.
- Most of existing methods are focused on qualitative evaluation of customers requirements. Methods for information quantification of customers preferences on product specifications are required for customers preferences segmentation and their optimal design specifications identification.
- Customers preferences information distortion due to inaccurate mapping of customers needs into design specifications, which always leading to improper decisions of specification combinations.

In this context, big sales data consist of comprehensive and multi-dimensional datasets including design features, functions, materials, performance, price ranges and their correspondent number of sales, etc. Such datasets are valuable sources for information mining of customers preferences on product specification combinations. In the era of Internet and E-commence, product data including sales and associated key specification combinations can be easily collected. Such data availability represents an emerging resource for supporting product design and development [4,12]. This paper aims at proposing a method to quantify customers preferences information of product specifications using big sales data. Segmental preferences of specification combinations are determined through density clustering with considerations of similarities among product specifications, and the best specification combinations of each segmental preferences cluster are obtained through cluster analysis.

2. Proposed method

In this section, a method to quantify customers preferences information is provided first. Then, customers preferences information mining is conducted by using big sales data. Design recommendations are provided for supporting design decision of product specifications by analyzing the obtained customers preferences information. A flowchart diagram of the proposed method is shown in Fig. 1.

2.1. Information quantification of customers preferences on product specifications

Customers requirements of products can be described by a set of specifications. Suppose *n* specifications are considered

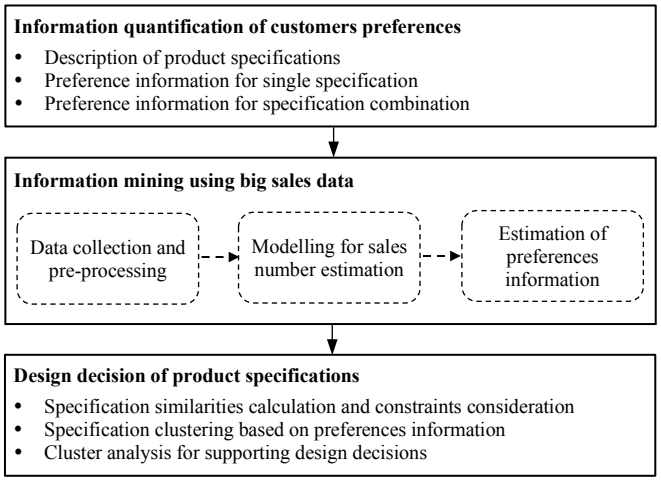

Figure 1. Diagram of the proposed method

for a category of products, the product specifications set $\boldsymbol{S}$ can be described as per Eq. 1:

$$\boldsymbol{S} = [S_1, S_2, \dots, S_n] \tag{1}$$

To simplify the discussion, discrete values of each specification are considered. According to Shannon's information entropy [13], the information entropy of customers demands on a specification value can be calculated as per Eq. 2:

$$H(S_i) = -\sum_{x=1}^{m_i} P\left(V_{i,x}\right) \cdot \log_2 P\left(V_{i,x}\right) \tag{2}$$

where $V_{i,x}$ represents the *x*-th value of S_i, $P\left(V_{i,x}\right)$ represents the probability of S_i to be selected as the value $V_{i,x}$, m_i represents the number of values of S_i.

To fully describe the customers requirements of product specifications, values of each specifications should be determined. The customers preferences information of product specifications, $H(\boldsymbol{S})$, can be calculated according to Eq. 3:

$$H(\boldsymbol{S}) = H(S_1, S_2, \dots, S_n)$$

$$= -\sum_{x=1}^{m_1}\sum_{y=1}^{m_2}\cdots\sum_{z=1}^{m_n} P\left(V_{1,x}, V_{2,y}, \dots, V_{n,z}\right) \cdot \log_2 P\left(V_{1,x}, V_{2,y}, \dots, V_{n,z}\right) \tag{3}$$

where $V_{1,x}, V_{2,y}, \dots, V_{n,z}$ represent the *x*-th, *y*-th, and *z*-th values of $S_1, S_2, \dots, S_n$, respectively. $m_1, m_2,$ and m_n represent the number of discrete values of $S_1, S_2, \dots, S_n$, respectively. $P\left(V_{1,x}, V_{2,y}, \dots, V_{n,z}\right)$ is the joint probability of $V_{1,x}, V_{2,y}, \dots, V_{n,z}$.

According to the information entropy theory, when the probability of each specification is equal to each other, and specifications are independent with each other, the joint information entropy of the specification combination can be obtained as per Eq. 4:

$$H(\boldsymbol{S})_{max} = -\log_2 \prod_{x=1}^{n} m_x \tag{4}$$

where n is the number of specifications, $m_1, m_2,$ and m_n represent the number of discrete values of $S_1, S_2, \dots, S_n$, respectively.

In general, the larger information entropy, the lower uncertainties on customers preferences of product specifications. To determine the values of product

specifications, a large quantity customers preferences information of product specifications, $H(\boldsymbol{S})$, is required to be obtained.

2.2. Information mining of customers preferences using product sales data

Customers preferences of product specifications are influencing in the purchasing behaviours. In this section, big data of product sales are used for the information mining of customers preferences on product specifications. In the proposed method, data are collected and processed for the calculation of probability density per specification combinations of products. An intelligent regression method [4,12] is adopted to predict probability density for a new set of product specifications. Major operations in the developed method are described as follows.

(a) Data collection and pre-processing

In this step, raw data of product specifications are collected for a given group of products through market survey. The survey output is a catalogue of products available on the market, including price range, number of monthly sales and a comprehensive list of product specifications. Following data acquisition, a data preparation procedure needs to be carried out. A first screening of the non-relevant specifications can be manually done taking into account the scope and the objective of the investigation. Consequently, the redundant specifications should be removed. Furthermore, a dimensionality reduction operation may be performed by removing those entries (product instances and specifications) showing a non-acceptable amount of missing data [14]. The result of this step consists in a complete matrix dataset. The big sales dataset consists of a large matrix with rows representing individual product instances and the columns representing the specifications along with sales number.

(b) Intelligent modelling for sales number estimation

Product sales data are influenced by product specifications and their combinations. The objective of this step is to build the relationships between product specifications and product sales number based on the available dataset. In order to build a specifications-sales number model, a multivariate regression operation should be carried out on the big dataset to model the relationship between the specifications and the sales number. Supervised learning techniques for intelligent regression can be used in this phase, where the input dataset is represented by the instances-specifications matrix and the target dataset is represented by the sales vector. The result of the intelligent regression consists in estimating the sales for each feasible combination of specifications values.

(c) Estimation of customers preferences information

The objective of this operation is to estimate the expected probability of market preferences of any specification combinations based on the available dataset. In this work, the total number of product sales N_T is calculated as follows:

$$N_T = \sum_{i=1}^{T} N_i \quad (5)$$

where T is the number of feasible specification combinations considered in testing. N_i is the estimated sales number of the i-th product, $i = 1, 2, \ldots, T$.

The probability of the market or customers' selection of the specification combination as $V_{1,i}, V_{2,i}, \ldots, V_{n,i}$, $P(V_{1,i}, V_{2,i}, \ldots, V_{n,i})$, can be calculated as:

$$P(V_{1,x}, V_{2,y}, \ldots, V_{n,z}) = N(V_{1,x}, V_{2,y}, \ldots, V_{n,z})/N_T \times 100\% \quad (6)$$

where $N(V_{1,x}, V_{2,y}, \ldots, V_{n,z})$ represents the estimated sales number of products with which the values of specifications were selected as $V_{1,x}, V_{2,y}, \ldots, V_{n,z}$.

In this work, all the possible specification combinations are considered for the testing of information estimation. By using Equations (3) and (6), information entropy of each possible product specifications can be calculated for supporting design decisions.

2.3. Design decision on product specifications

Products with similar features are considered for avoiding both intensive competitions among those products and unnecessary resource consumption. In this work, similarities of product specifications are calculated first. Then clustering of specifications by considering both specifications similarities and customers preferences information on specifications is carried out for customers preferences segmentation. At last, predominance analysis of specifications of each cluster is carried out to identify the optimal specification combinations.

2.3.1 Specification similarities calculation and physical constraints consideration

Customers product preferences usually change with values of product specifications. The higher similarities of values of product specifications, the lower differences of customers preferences on those products. To avoid extensive internal competition amongst newly designed products, similarities among product specifications is calculated for customers preferences segmentation.

Calculation of similarities between two vectors has been well studied and various metrics (such as Euclidean distance, Manhattan distance, Cosine similarity, Spearman Rank Correlation) can be adopted for the selection depending on different engineering needs. Since the units and scales of each specifications may differ from each other, in order to balance the influence of different specifications on customers preferences, a data normalization procedure needs to be carried out for all the specification values.

In this work, suppose $V_{1,x}, V_{2,y}, \ldots, V_{n,z}$ represent the x-th, y-th, and z-th values of $S_1, S_2, \ldots, S_n$, and $V_{1,x'}, V_{2,y}, \ldots, V_{n,z'}$ represent the x'-th, y'-th, and z'-th values of $S_1, S_2, \ldots, S_n$, respectively, then the Manhattan similarity between specification array $SV = [V_{1,x}, V_{2,y}, \ldots, V_{n,z}]$ and $SV' = [V_{1,x'}, V_{2,y}, \ldots, V_{n,z'}]$ can be calculated as follows:

$$\text{Similarity } (SV, SV') = |V_{1,x} - V_{1,x'}| + |V_{2,y} - V_{2,y'}| + \cdots + |V_{n,z} - V_{n,z'}| \quad (7)$$

In addition to specification similarity calculation, constraints among product specifications also need to be defined for the physical feasibility consideration. In this work, constraints of product specifications are defined as follows:

$$f_u(S_1, S_2, \dots, S_n) = 0, g_v(S_1, S_2, \dots, S_n) \leq 0, u, v = 1, 2, \dots \quad (8)$$

where $f_u()$ and $g_v()$ represents the relationship functions among the specifications.

Specification combinations that don't meet the constraints defined in Equation (8) are not considered for similarity calculation and specification clustering.

2.3.2 Specification clustering based on preferences information

Similarities and constraints among product specifications are used for specification clustering based on preferences information for customers preferences segmentation. The objective of customers preferences segmentation is to divide the market into different segments based on customers preferences of product specifications. With reference to the customers preferences information obtained in the previous section, in order to allow for design support, specification clustering is carried out with consideration of both product similarities and information entropy for market segmentation.

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) is used to automatically cluster samples into different groups based on their distribution characteristic [15]. To use DBSCAN, two thresholds are defined in this work: a similarity threshold S_T and a preferences information threshold P_T.

The similarity threshold S_T is defined for assessing the similarity between each pair of specification combinations within the same cluster. Calculation of similarity between each pair of specification combinations was provided in previous section. Setting a larger similarity threshold S_T would yield to a higher similarity of product specifications within the same cluster.

The preferences information threshold P_T is defined for the consideration of the degree of customers preferences which can be quantified by the information entropy $H(\boldsymbol{S})$ defined in Equation (3). The preferences information threshold P_T is set up for allowing designs (specification combinations) in the same cluster to be with higher preferences information (higher than P_T). Determination of thresholds P_T and S_T depends on the designer expertise.

Suppose a design with two specifications S_1 and S_2, as shown in Figure 2, the preferences information of each feasible design appears as a highly nonlinear function. The preferences information threshold P_T determines the number of clusters. The similarity threshold S_T restricts the maximal similarities of product specifications of each cluster.

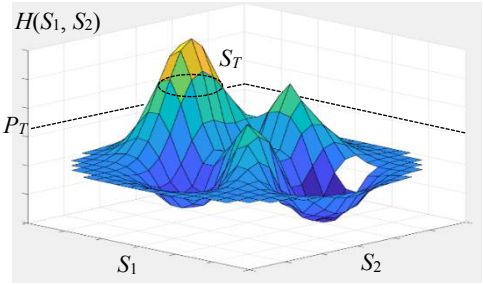

Figure 2. Segmentation of product specifications

Suppose c is the number of clusters, the segments of customers preferences ($\boldsymbol{SP}$) of product specifications in the market can be modelled as follows:

$$\boldsymbol{SP}=\{SP_1, SP_2, \dots, SP_c\} \quad (9)$$

Given that specification combinations of values in each cluster are characterised with an acceptable preferences information (higher than the predefined threshold P_T), at least one specification combination of values should be selected as design candidate. While specification combinations of values in the same clusters are approximately similar to each other (higher than the predefined similarity threshold S_T), in order to avoid undesired internal competitions, cluster analysis needs to be carried out to identify the predominant specification combinations of values for supporting design decisions.

2.3.3 Cluster analysis for supporting design decisions

As mentioned in the previous section, the objective of the cluster analysis is to assess the predominance of specification value combinations within each cluster. Based on the results analysis and interpretation, the clusters in this step can be divided into two categories, i.e. single design candidate clusters and multiple design candidate clusters according to the number of predominant combinations of specification values contained in each cluster. If a statistical descriptor, i.e. the kurtosis, is larger than a predefined threshold value, then a design candidate value combination of specifications can be considered as "predominant" [12]. The threshold value can be empirically defined by designers/experts based on their experience.

Following this subdivision, a statistical test needs to be carried out with the objective of verifying whether there exists a predominant value (or combinations of values) among those design features yielding higher information entropy.

Scenario A: there exists a predominant combination of values. In this case, the predominant combination of values can be defined as target specification values. Original Equipment Manufacturer (OEM) can apply the combinations of values to a design without change.

Scenario B: there exists multiple predominant specification combinations of values. In such scenario, a set of products or product family should be built by OEMs based on these specification values, to meet the customers needs.

3. Case Study

Electric bicycle is selected as a product category to illustrate the proposed method. Monthly sales number and specification list of electric bicycles are collected from the shopping platform Alibaba© through keywords searching. From the specifications list, a dataset reduction procedure was applied to remove irrelevant and redundant specifications. A statistical characterization was conducted to identify missing data for each product and for each specification. Missing data over 40% yielded to the entry removal. This procedure reduced the dataset size to 133 electric bicycles × 12 specifications. Specifications and their values are reported in Table 1.

In addition to the value ranges of each specification reported in Table 1, three additional constraints of specifications were defined in this case study as follows:

- Weight = *'(50, 80]'* → Foldability = *'Un-foldable'*
- Weight = *'(5, 10]'* → Passenger number = *'1'*
- Price = *'(900, 1400]'* → Endurance = *'(25, 35]'* or *'(35, 45]'*

In this work, the regression model was obtained using

Neural Network for data fitting. Input dataset was made of 133×12 matrix, where the 133 rows represent the number of product instances and the 12 columns represent the number of specifications. A z-score [16] normalization procedure was applied to all the specification values in the input dataset. The target dataset consisted in a 133 elements vector containing the number of sales corresponding to each bicycle instance. The training and testing procedure required the partition of the initial input-output vectors into training, validation and test subsets according to the following percentages: training subset = 70%; validation subset = 15%; testing subset = 15%. The most accurate configuration resulted to be 30 hidden layer nodes and Bayesian Regularization training algorithm with a high coefficient of determination 0.923, which minimizes a linear combination of squared errors and weights, providing for good generalization capabilities.

(a)

Specifications	Values
Max speed (Km/h)	(20, 25]
Endurance (Km)	(25, 35]
Battery type (-)	Lead-acid
Price (CNY)	1500
Length (mm)	1470
Height (mm)	1052
…	…

(b)

Figure 3. (a) Electric bicycle specimen and (b) its specifications

Table 1. Specifications and their value ranges

ID	Specification	Value	Unit
1.	Max Speed	(5-20]; (20, 25]; (25, 30]; (30, 40]; (40, 60]	Km/h
2.	Passenger number	1; 2	-
3.	Battery type	Lithium-iron; Lead-acid; Graphene	-
4.	Weight	(5, 10]; (10, 15]; (15, 25]; (25, 50]; (50, 80]	Kg
5.	Max load	(50, 100]; (100, 150]; (150, 200]; (200, 300]	Kg
6.	Material	High carbon steel; Aluminum alloy; Magnesium alloy	-
7.	Foldability	Un-foldable; Foldable	-
8.	Rim size	<14; =14; >14	Inch
9.	Shock absorber (SA)	Without shock absorber; Fork shock absorber; Fork SA/ Middle SA; Fork SA/Rear spring absorber; Fork-hydraulic SA /Middle SA/Front SA; Middle SA; Middle-rear SA; Rear absorber; Seat SA; Fork-spring SA /Seat SA; Front SA/Middle SA/Seat SA; Fork-hydraulic SA /Seat SA/Front SA; Fork-hydraulic SA /Seat SA/Middle SA/Front SA	-
10.	Brake mode	Disc brakes; Drum brakes; Front drum & back expansile brakes; Front drum & back disc brakes; Front disc & back drum brakes; Front disc & back expansile brakes; Front V & back roller brakes; Front drum & back servo brakes; Front V & back servo brakes; Front drum & back roller brakes; Front disc & back V brakes	-
11.	Endurance	(25, 35]; (35, 45]; (45, 55]; (55, 65]; (65, 100]	Km
12.	Price	(900, 1400]; (1400, -1800]: (1800, 2200]; (2200, 2650]; (2650, 3050]; (3050, 3500]; (3500, -3900]; (3900, 4400]; (4400, 4800]: (4800, 5200]; (5200, 6000]; (6000, 10000]	CNY

To estimate the number of sales of new specification combinations, the trained neural network has been then tested on a new dataset. Such dataset was obtained calculating all the possible combinations of all the values of specifications reported in Table 1. Consequently, those combinations which do not satisfy the above-mentioned feasibility constraints were removed from the dataset. The results, i.e. the estimated number of sales produced by the neural network tested on such a large dataset were then sorted in descending order and the combinations yielding negative values of sales were discarded.

By using Equations (3) and (6), information entropy of each specification combination can be calculated for clustering analysis. It is realized that homogeneous competition between products with similar specification combinations needs to be avoided for reducing unnecessary resource consumption. In this work, similarities between each pair of specification combinations are calculated by using Equation (7).

Specification clustering based on preferences information with consideration of product similarity is conducted by setting the preferences information threshold P_T to 0.25, and similarity threshold S_T to 0.8. In this step, 3 segmental preferences of product specifications were identified. Through frequency analysis of each specification cluster, predominant combinations of specification values of clusters were obtained as summarized in Table 2. It is estimated that the optimal design specification combinations summarized in Table 2 are with high desired preference of customers and undesired mutual competitions of these design candidates can be avoided.

Table 2 Optimal specification combinations

Clusters	Specification combinations
Cluster 1	{Max Speed: (20, 25]; Passenger number: 2; Battery type: Lead-acid Weight: (15, 25]; Max load: (50, 100]; Material: High carbon steel; Foldability: Un-foldable; Rim size: =14; Shock absorber: Fork SA/Rear spring absorber; Blake mode: Front drum & back expansile brakes; Endurance: (65, 100]}
	{Max Speed: (20, 25]; Passenger number:1; Battery type: Lead-acid Weight: (15, 25]; Max load: (50, 100]; Material: High carbon steel; Foldability: Un-foldable; Rim size: =14; Shock absorber: Fork SA/Rear spring absorber; Blake mode: Front drum & back expansile brakes; Endurance: (55, 65]}
	{Max Speed: (5-20]; Passenger number: 2; Battery type: Lead-acid; Weight: (15, 25]; Max load: (50, 100]; Material: High carbon steel; Foldability: Un-foldable; Rim size: =14; Shock absorber: Fork SA/Rear spring absorber; Blake mode: Front drum & back expansile brakes; Endurance: (55, 65]}
Cluster 2	{Max Speed: (20, 25]; Passenger number: 2; Battery type: Lithium-iron Weight: (15, 25]; Max load: (150, 200]; Material: High carbon steel; Foldability: Foldable; Rim size: =14; Shock absorber: Fork SA/Rear spring absorber; Blake mode: Disc brakes; Endurance: (55, 65]}
Cluster 3	{Max Speed: (20, 25]; Passenger number: 2; Battery type: Lithium-iron; Weight: (15, 25]; Max load: (100, 150]; Material: High carbon steel; Foldability: Foldable; Rim size: =14; Shock absorber: Fork SA/Rear spring absorber; Blake mode: Front drum & back expansile brakes; Endurance: (55, 65]}
	{Max Speed: (25, 30]; Passenger number: 2; Battery type: Lithium-iron; Weight: (15, 25]; Max load: (150, 200]; Material: High carbon steel; Foldability: Foldable; Rim size: =14; Shock absorber: Fork SA/Rear spring absorber; Blake mode: Front drum & back expansile brakes; Endurance: (55, 65]}

4. Discussions and conclusions

4.1. Discussions

Although the newly developed method has some advantages for competitive product specifications determination, the following aspects need to be carefully treated as the accuracy of results is highly sensitive to such factors: (1) data collection and pre-processing for intelligent relation training, (2) method for specifications similarity calculation, and (3) selections of similarity threshold S_T and preferences information threshold P_T for specification clustering. In addition, the application scopes of the proposed method are limited to maturely developed products and to ensured quality of data collection.

Nevertheless, emerging big sales data provide for valuable resources for supporting design-decision of competitive product specifications. In addition to information mining of customers preferences, the following issues need to be investigated:

- Product evolution analysis for design adaptations using big sales data. Products constantly evolve through mutation, crossover and reproduction of products specifications in the marketplace to better satisfy customers preferences in the market. To better support design adaptations, methods are needed for product evolution analysis using big historical sales data.
- Game analysis of product specifications (GAPS) for enhancing product competitiveness. Generally, products with similar specification combinations usually compete directly with each other, especially when customers make comparisons before making their purchases decisions. In order to support rationalized design decisions, game analysis of product specifications need to be carried out for the investigations of product competition mechanism.
- Correlation analysis of product specifications (CAPS) for supporting modular architecture design. Correlations among specifications originated from customer conscious and subconscious preferences are embedded in sales data. In addition to existing methods tend to use independence and/or similarity based on considerations of component geometry, materials, assembly and disassembly, etc, for product modularity. The unambiguous functional correlation analysis among components remains the less regarded topics in product modularization. It was observed that prediction and incorporation of the diversified market preferences on product specification combinations are rarely studied for supporting modular design decision.

4.2. Conclusions

In order to improve product competitiveness in the market, a novel method is developed for design specifications determination based on information mining of customers preferences using big sales data. In the proposed method, information entropy is used to quantify the customers preferences of product specifications. The preferences information of customers is estimated by using collected sales data. To determine the optimal product specifications with consideration of segmental preferences in the market, a density-based clustering analysis is carried out. Electric bicycle specifications determination using big sales data is reported to illustrate and validate the proposed method.

Despite the progress reported in this work, more research and application efforts are needed to further facilitate product design using big data of product sales. Ongoing research activities are being carried out on:

- Product evolution analysis for design adaptation using big historical data of product sales.
- Game analysis of product specifications for enhancing product competitiveness.
- Design of adaptable product architecture and interfaces considering customers preferences information.

Acknowledgments

The authors wish to thank the National Key R&D Program of China (No: 2018YFB1701701) for providing financial support to this research.

References

[1] Jiao J and Chen CH. 2006. Customer requirement management in product developmnet: a review of research issues. Concurrent Engineering: Research and Applications: 14(3):173-185.
[2] Ireland R, Liu A. 2018. Application of data analytics for product design: sentiment analysis of online proudct reviews. CIRP Journal of Manufacturing Science and Technology, 23:128-144.
[3] Ulrich KT, Eppinger SD. 2004. Product design and development. McGraw-Hill Education, New York.
[4] Zhang J, Simeone A, Gu P, Hong B. 2018. Product features characterization and customers' preferences prediction based on purchasing data. CIRP Annals – Manufacturing Technology, 67(1):149-152.
[5] Maslow A.H. 1987. Motivation and personality (3rd ed.). Pearson Education, Delhi.
[6] Kano N, Seracu N, Takahashi F, Tsuji S. 1984. Attractive quality and must-be quality. The Journal of Japanese Society for Quality Control, 14(2):147-152.
[7] Jin J, Liu Y, Ji P, Kwong CK. 2019. Review on recent advances in information mining from big consumer opinion data for product design. Journal of Computing and Information Science Engineering, 19(1):010801
[8] Cherif MS, Chabchoub H, Aouni B. 2010. Integrating customer's preferences in the QFD planning process using a combined benchmarking and imprecise goal programming model. Internationl Transactions in Operational Research. 17(1):85-102.
[9] Yan H B , Ma T . 2015. A group decision-making approach to uncertain quality function deployment based on fuzzy preference relation and fuzzy majority. European Journal of Operational Research, 241(3):815-829.
[10] Zhao H, Liu Q, Ge Y, Kong R, Chen E. 2016. Group preference aggregation: a Nash equilibrium approach. Transactions of the 16th IEEE International Conference on Data Mining Workshops, December 12-15, Barcelona, Spain.
[11] Liu W, Wang Y. 2022. Research on the spatial optimal aggregation method of decision maker preference information based on Steiner-Weber point. Computers & Industrial Engineering, 163:107819.
[12] Zhang J, Chu X, Simeone A, Gu P. 2021. Machine learning-based design feature decision support tool via customers purchasing data analysis. Concurrent Engineering: Research and Applications, 29(2):124-141.
[13] Shannon CE. 1948. A mathematical theory of communication. Bell System Technical Journal, 27(3):379-423.
[14] Kramer O. 2013. On missing data hybridizations for dimensionality reduction. In: Hybrid Metaheuristics (eds MJ Blesa, C Blum, P Festa, et al.), pp. 189-197. Springer, Berlin Heidelberg.
[15] Ester M, Kriegel H-P, Sander J, Xu X. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, August 2-4, Portland, Oregon.
[16] Nawi NM, Atomi WH, Rehman MZ. 2013. The effect of data pre-processing on optimized training of artificial neural networks. Procedia Technology, 11:32-39.